

IN THE CLAIMS:

1. (Currently amended) A method of selecting data sets for use with a predictive algorithm of customer behavior, comprising:
 - generating a first geographical distribution of a training data set for a predictive algorithm of customer behavior, said training data set being derived from a database containing customer information;
 - generating a second geographical distribution of a testing data set for said predictive algorithm of customer behavior, said testing data set being derived from said database containing customer information;
 - comparing the first geographical distribution and the second geographical distribution to identify a discrepancy between the first geographical distribution and the second geographical distribution; and
 - modifying selection of entries in one or more of the training data set and the testing data set based on the discrepancy between the first geographical distribution and the second geographical distribution.
2. (Currently amended) The method of claim 1, wherein the first geographical distribution and the second geographical distribution are distributions of drive time from a customer geographical location to a commercial establishment geographical location.
3. (Currently amended) The method of claim 1, wherein the first geographical distribution and the second geographical distribution are distributions of distance between a customer geographical location and a commercial establishment geographical location.
4. (Currently amended) The method of claim 1, wherein comparing the first geographical distribution and the second geographical distribution includes comparing one or more of a mean, mode, and standard deviation of the first geographical distribution to one or more of a mean, mode, and standard deviation of the second geographical distribution.

5. (Currently amended) The method of claim 1, wherein the first geographical distribution and the second geographical distribution are distributions of a weighted distance between a customer geographical location and commercial establishment geographical locations.
6. (Currently amended) The method of claim 1, wherein the first geographical distribution and the second geographical distribution are distributions of a weighted drive time between a customer geographical location and commercial establishment geographical locations.
7. (Original) The method of claim 1, wherein modifying selection of entries in one or more of the training data set and the testing data set includes generating recommendations for improving selection of entries in one or more of the training data set and the testing data set.
8. (Currently amended) The method of claim 1, ~~wherein the~~ further comprising the step of selecting said training data set and [[the]] said testing data set ~~are selected~~ from a customer information database.
9. (Currently amended) The method of claim 1, further comprising comparing at least one of the first geographical distribution and the second geographical distribution to a geographical distribution of a customer database.
10. (Currently amended) The method of claim 1, wherein the first geographical distribution and second geographical distribution are frequency distributions of one of drive time and distance between a customer geographical location and one or more commercial establishment geographical locations.
11. (Currently amended) The method of claim 9, wherein comparing at least one of the first geographical distribution and the second geographical distribution to a geographical distribution of a customer database includes:

generating a composite data set from the training data set and the testing data set;
and

generating a composite geographical distribution from the composite data set.

12. (Original) The method of claim 1, wherein modifying selection of entries in one or more of the training data set and the testing data set includes changing one of a random selection algorithm and a seed value for a random selection algorithm.

13. (Currently amended) The method of claim 1, further comprising training ~~[[a]]~~ said predictive algorithm for customer behavior using at least one of the training data set and the testing data set if the discrepancy is within a predetermined tolerance.

14. (Currently amended) The method of claim 13, wherein the predictive algorithm for customer behavior is a discovery based data mining algorithm.

15. (Currently amended) An apparatus for selecting data sets for use with a predictive algorithm of customer behavior, comprising:

a statistical engine; and

a comparison engine coupled to the statistical engine, wherein the statistical engine generates a first geographical distribution of a training data set of customer information and a second geographical distribution of a testing data set of customer information, the comparison engine compares the first geographical distribution and the second geographical distribution to identify a discrepancy between the first geographical distribution and the second geographical distribution and modifies selection of entries in one or more of the training data set and the testing data set based on the discrepancy between the first geographical distribution and the second geographical distribution.

16. (Currently amended) The apparatus of claim 15, wherein the first geographical distribution and the second geographical distribution are distributions of drive time from a customer geographical location to a commercial establishment geographical location.

17. (Currently amended) The apparatus of claim 15, wherein the first geographical distribution and the second geographical distribution are distributions of distance between a customer geographical location and a commercial establishment geographical location.

18. (Currently amended) The apparatus of claim 15, wherein the comparison engine compares the first geographical distribution and the second geographical distribution by comparing one or more of a mean, mode, and standard deviation of the first geographical distribution to one or more of a mean, mode, and standard deviation of the second geographical distribution.

19. (Currently amended) The apparatus of claim 15, wherein the first geographical distribution and the second geographical distribution are distributions of a weighted distance between a customer geographical location and commercial establishment geographical locations.

20. (Currently amended) The apparatus of claim 15, wherein the first geographical distribution and the second geographical distribution are distributions of a weighted drive time between a customer geographical location and commercial establishment geographical locations.

21. (Original) The apparatus of claim 15, wherein the comparison engine modifies selection of entries in one or more of the training data set and the testing data set by generating recommendations for improving selection of entries in one or more of the training data set and the testing data set.

22. (Original) The apparatus of claim 15, further comprising a training data set/testing data set selection device that selects the training data set and the testing data set from a customer information database.

23. (Currently amended) The apparatus of claim 15, wherein the comparison engine further compares at least one of the first geographical distribution and the second geographical distribution to a geographical distribution of a customer database.

24. (Currently amended) The apparatus of claim 15, wherein the first geographical distribution and second geographical distribution are frequency distributions of one of drive time and distance between a customer geographical location and one or more commercial establishment geographical locations.

25. (Currently amended) The apparatus of claim 23, wherein the comparison engine compares at least one of the first geographical distribution and the second geographical distribution to a geographical distribution of a customer database by:

generating a composite data set from the training data set and the testing data set;
and

generating a composite geographical distribution from the composite data set.

26. (Original) The apparatus of claim 15, wherein the comparison engine modifies selection of entries in one or more of the training data set and the testing data set by changing one of a random selection algorithm and a seed value for a random selection algorithm.

27. (Original) The apparatus of claim 15, further comprising a predictive algorithm device, wherein the predictive algorithm device is trained using at least one of the training data set and the testing data set if the discrepancy is within a predetermined tolerance.

28. (Original) The apparatus of claim 27, wherein the predictive algorithm is a discovery based data mining algorithm.

29. (Currently amended) A computer program product in a computer readable medium for selecting data sets for use with a predictive algorithm of customer behavior, comprising:

first instructions for generating a first geographical distribution of a training data set derived from customer information;

second instructions for generating a second geographical distribution of a testing data set derived from customer information;

third instructions for comparing the first geographical distribution and the second geographical distribution to identify a discrepancy between the first geographical distribution and the second geographical distribution; and

fourth instructions for modifying selection of entries in one or more of the training data set and the testing data set based on the discrepancy between the first geographical distribution and the second geographical distribution.

30. (Currently amended) The computer program product of claim 29, wherein the first geographical distribution and the second geographical distribution are distributions of drive time from a customer geographical location to a commercial establishment geographical location.

31. (Currently amended) The computer program product of claim 29, wherein the first geographical distribution and the second geographical distribution are distributions of distance between a customer geographical location and a commercial establishment geographical location.

32. (Currently amended) The computer program product of claim 29, wherein the third instructions for comparing the first geographical distribution and the second geographical distribution include instructions for comparing one or more of a mean, mode, and standard deviation of the first geographical distribution to one or more of a mean, mode, and standard deviation of the second geographical distribution.

33. (Currently amended) The computer program product of claim 29, wherein the first geographical distribution and the second geographical distribution are distributions of a weighted distance between a customer geographical location and commercial establishment geographical locations.

34. (Currently amended) The computer program product of claim 29, wherein the first geographical distribution and the second geographical distribution are distributions of a weighted drive time between a customer geographical location and commercial establishment geographical locations.

35. (Original) The computer program product of claim 29, wherein the fourth instructions for modifying selection of entries in one or more of the training data set and the testing data set include instructions for generating recommendations for improving selection of entries in one or more of the training data set and the testing data set.

36. (Currently amended) The computer program product of claim 29, further comprising fifth instructions for comparing at least one of the first geographical distribution and the second geographical distribution to a geographical distribution of a customer database.

37. (Currently amended) The computer program product of claim 29, wherein the first geographical distribution and second geographical distribution are frequency distributions of one of drive time and distance between a customer geographical location and one or more commercial establishment geographical locations.

38. (Currently amended) The method of claim 36, wherein the fifth instructions include:

instructions for generating a composite data set from the training data set and the testing data set; and

instructions for generating a composite geographical distribution from the composite data set.

39. (Original) The computer program product of claim 29, wherein the fourth instructions for modifying selection of entries in one or more of the training data set and the testing data set include instructions for changing one of a random selection algorithm and a seed value for a random selection algorithm.

40. (Currently amended) The computer program product of claim 29, further comprising fifth instructions for training a predictive algorithm of customer behavior using at least one of the training data set and the testing data set if the discrepancy is within a predetermined tolerance.